Такой MySQL мне нужен самому,

или Как мы докатились до multisource

Константин Рекунов





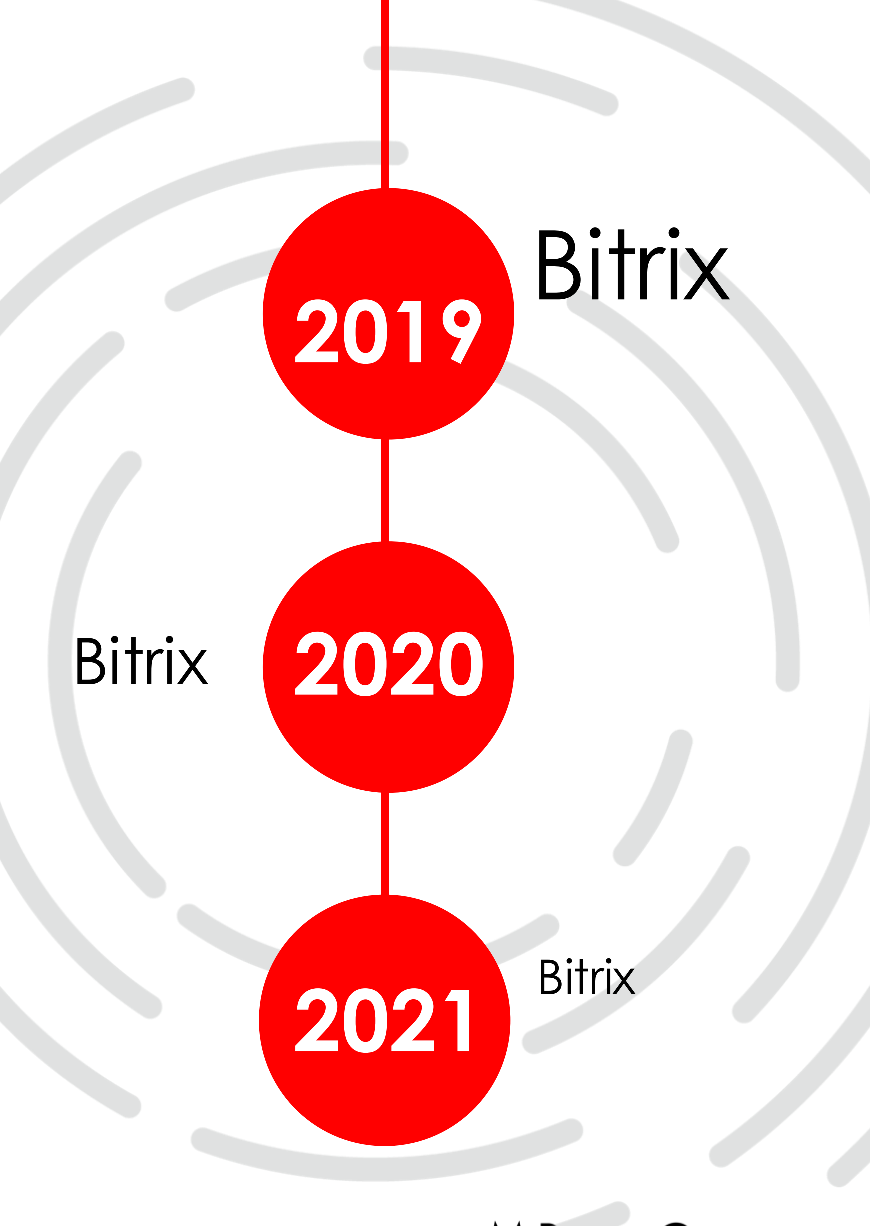
Предыстория



ПИЛИТЕ, ШУРА, ОНА ЗОЛОТАЯ

«Если ваша компания не разносила монолит, значит она не росла»

© Ревяшко А.С.



#ВКЛЮЧАЙБУДУЩЕЕ

Legacy Bitrix +

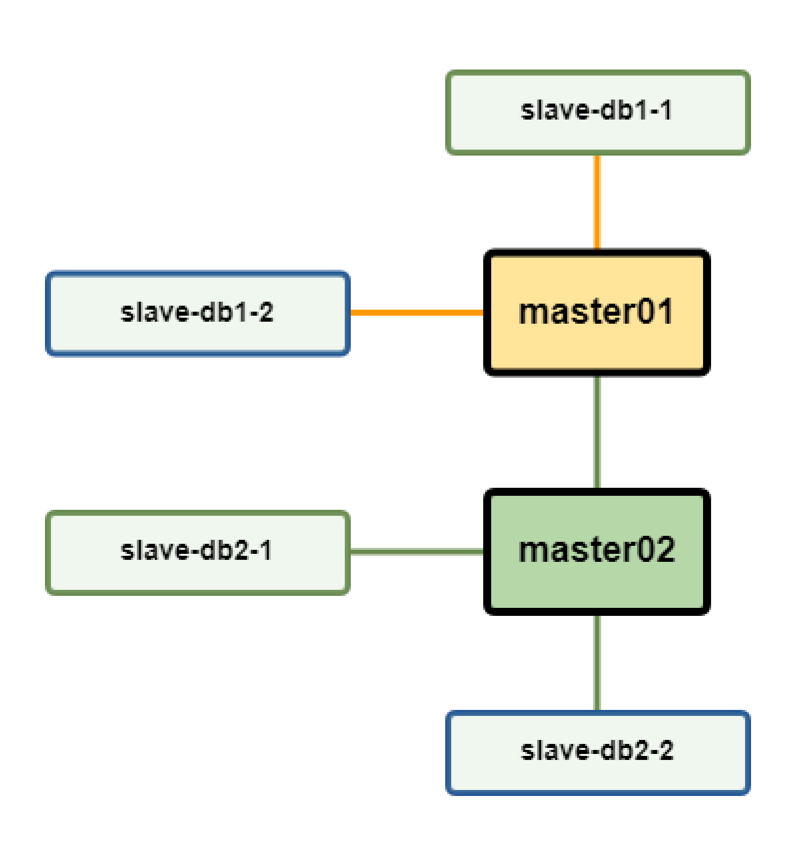
K8S, GoLang, NodeJS, JAVA

Percona MySQL

М.ВидеоЭльдорадо

- 1000+ SQL запросов на странице без кеша
- 40-50k общий rps на кластер БД
- Чтение с мастеров!

2 активных мастера?





М.ВидеоЭльдорадо



ВНИМАНИЕ

Все трюки выполнены профессионалами, ради собственной безопасности и безопасности окружающих не пытайтесь их повторить

Больше мастеров – больше стабильности?

slave-db1-1

slave-db1-2

master01

master02

slave-db2-1

slave-db2-2

slave-db4-1

master04

slave-db4-2

master03

slave-db3-1

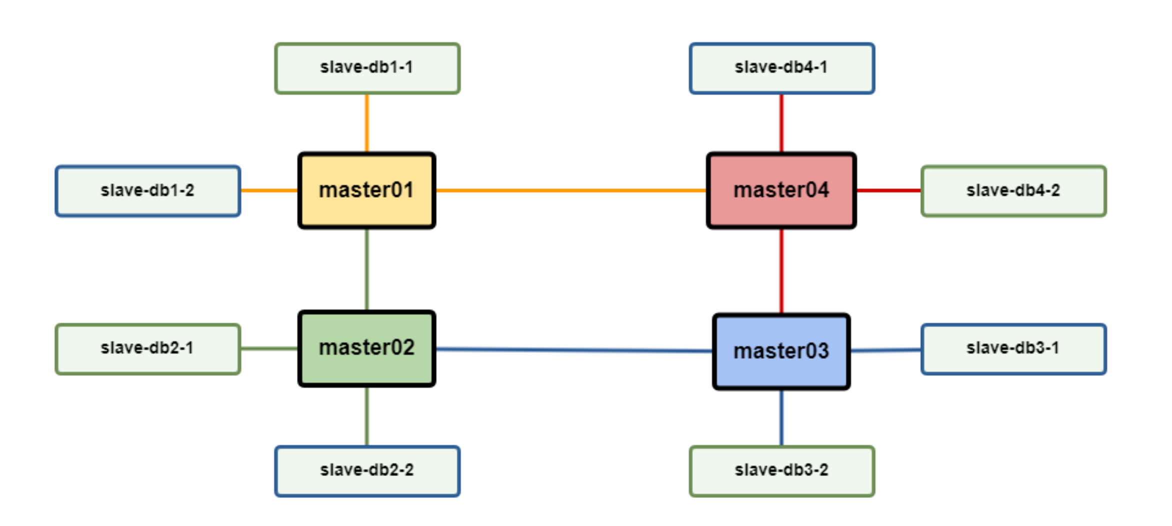
slave-db3-2

М.ВидеоЭльдорадо



Больше мастеров – больше стабильности?

- А если отказал любой из 4 master'ов?
- А если прилетел тяжелый апдейт?
- Сколько шагов репликации должен пройти апдейт с master01 до слейва 4-1?



М.ВидеоЭльдорадо



- 150 000 товаров * 100 регионов * цены
- 150 000 товаров * 100 регионов * остатки
- Логистические цепочки
- v5.5 на мастерах => binlog_row_image FULL

```
[/var/lib/mysql/mysql-bin-master02.001832] 2021-05-13 00:11:00 -- 2021-05-13 00:12:00
[/var/lib/mysql/mysql-bin-master02.001833] 2021-05-13 00:11:00 -- 2021-05-13 00:12:00
  15161 ### UPDATE `eldorado2`.`b_iblock_element_prop_s31`
  12000 ### DELETE FROM `eldorado2`.`mv_sap_group2type_transport`
  10371 ### UPDATE `eldorado2`.`b_catalog_product`
   3128 ### INSERT INTO `eldorado2`.`b_sale_fuser`
   1451 ### UPDATE `eldorado2`.`b_catalog_price`
00:12:00 2021-05-13 -- 00:13:00 2021-05-13
[/var/lib/mysql/mysql-bin-master02.001832] 2021-05-13 00:12:00 -- 2021-05-13 00:13:00
[/var/lib/mysql/mysql-bin-master02.001833] 2021-05-13 00:12:00 -- 2021-05-13 00:13:00
  36218 ### UPDATE `eldorado2`.`b_iblock_element_prop_s31`
   7000 ### DELETE FROM `eldorado2`.`mv_sap_group2type_transport`
   4877 ### UPDATE `eldorado2`.`b_catalog_product`
   2093 ### INSERT INTO `eldorado2`.`b_sale_fuser`
    175 ### UPDATE `eldorado2`.`b_catalog_price`
00:13:00 2021-05-13 -- 00:14:00 2021-05-13
[/var/lib/mysql/mysql-bin-master02.001832] 2021-05-13 00:13:00 -- 2021-05-13 00:14:00
[/var/lib/mysql/mysql-bin-master02.001833] 2021-05-13 00:13:00 -- 2021-05-13 00:14:00
  50002 ### DELETE FROM `eldorado2`.`b_sale_fuser`
  44609 ### UPDATE `eldorado2`.`b_iblock_element_prop_s31`
  12000 ### DELETE FROM `eldorado2`.`mv_sap_group2type_transport`
  11891 ### UPDATE `eldorado2`.`b_catalog_product`
   1667 ### UPDATE `eldorado2`.`b_catalog_price`
00:14:00 2021-05-13 -- 00:15:00 2021-05-13
[/var/lib/mysql/mysql-bin-master02.001832] 2021-05-13 00:14:00 -- 2021-05-13 00:15:00
[/var/lib/mysql/mysql-bin-master02.001833] 2021-05-13 00:14:00 -- 2021-05-13 00:15:00
  45257 ### DELETE FROM `eldorado2`.`b_sale_fuser`
  18862 ### UPDATE `eldorado2`.`b_iblock_element_prop_s31`
  12665 ### UPDATE `eldorado2`.`b_catalog_product`
  12000 ### DELETE FROM `eldorado2`.`mv_sap_group2type_transport`
   1712 ### UPDATE `eldorado2`.`b_catalog_price`
```

ЧТО ДЕЛАТЬ?

Синхронная репликация в БД медленная.

Асинхронная остается преобладающей для масштабирования скорости и доступности.





Совсем не в ту степь

GALERA CLUSTER

хорошо, но нет гарантии на blackhole

Percona XtraDB CLUSTER

галера под капотом? Нет гарантии на blackhole

НЕ ПОВТОРЯТЬ

Master: ALTER TABLE superbigtable ENGINE=BLACKHOLE;

Slave: ALTER TABLE superbigtable ENGINE=INNODB;

Ограничения

- ~~Ленивые админы~~
- Использование фишек репликации (blackhole обязательно)
- Максимальная совместимость с текущим кодом & инструментами обслуживания
- **Без даунтайма!**

М.ВидеоЭльдорадо

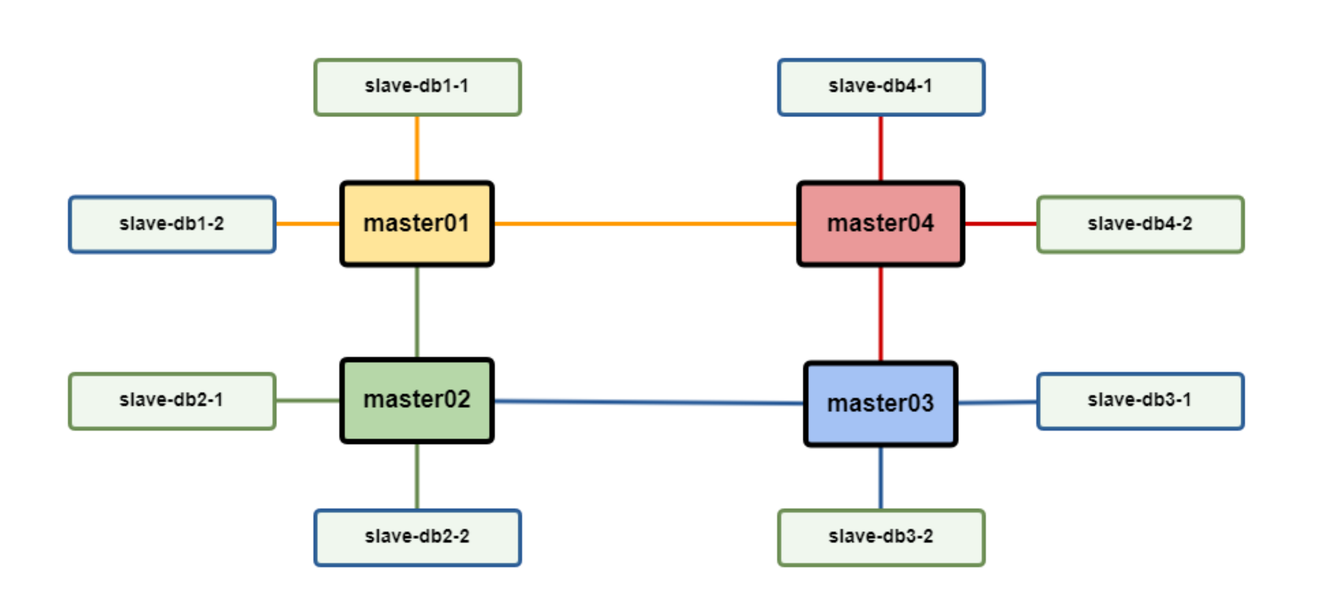


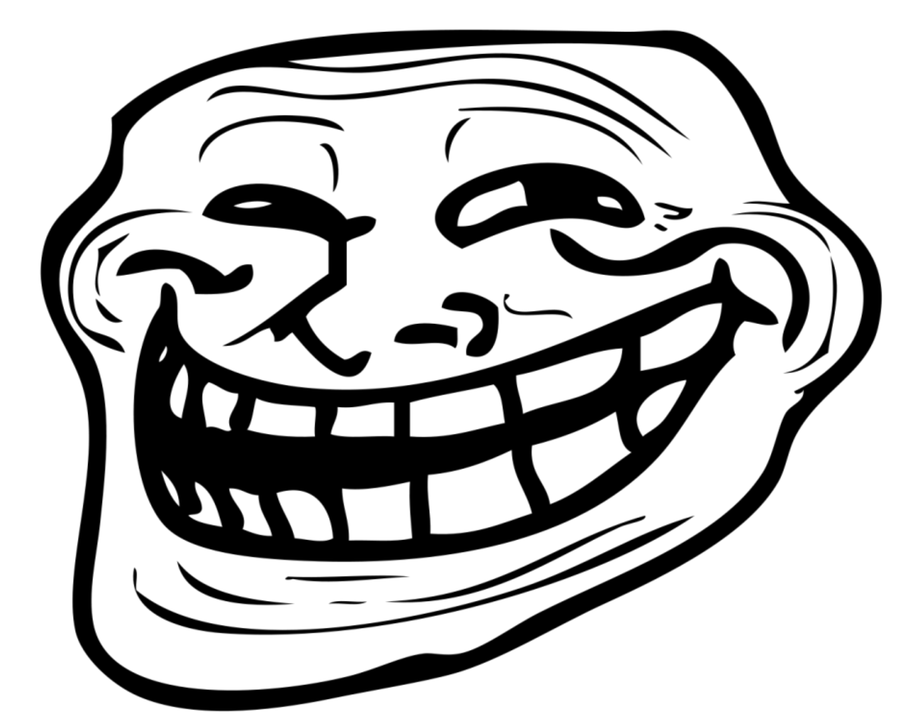
MySQL 5.5

5.5 на мастерах, 5.7 на слейвах

Официальная документация:

«when upgrading to MySQL 5.7, the slaves must be MySQL 5.7 before you can upgrade the master to 5.7.»





М.ВидеоЭльдорадо



Обновление с 5.5 до 5.7

binlog_checksum=NONE

binlog_row_image=FULL

binlog_rows_query_log_events=OFF

log_bin_use_v1_row_events=1



с 5.5 на 5.7 в кольце

```
mysql -h127.0.0.1 -P 3302 -e "RESET SLAVE; RESET MASTER; SHOW MASTER STATUS\G;"
mysql -h127.0.0.1 -P 3302 -e "CHANGE MASTER TO MASTER_HOST='127.0.0.1', MASTER_PORT=3301, MASTER_USER='root', MASTER_PASSWORD='', MASTER_LOG_FILE='master1-binlog.000001', MASTER_LOG_POS = 0;START SLAVE;"
mysql -h127.0.0.1 -P 3303 -e "RESET SLAVE; RESET MASTER; SHOW MASTER STATUS\G;"
mysql -h127.0.0.1 -P 3303 -e "CHANGE MASTER TO MASTER_HOST='127.0.0.1', MASTER_PORT=3302, MASTER_USER='root', MASTER_PASSWORD='', MASTER_LOG_FILE='master2-binlog.000001', MASTER_LOG_POS = 0;START SLAVE;"
mysql -h127.0.0.1 -P 3304 -e "RESET SLAVE; RESET MASTER; SHOW MASTER STATUS\G;"
mysql -h127.0.0.1 -P 3304 -e "CHANGE MASTER TO MASTER_HOST='127.0.0.1', MASTER_PORT=3303, MASTER_USER='root', MASTER_PASSWORD='', MASTER_LOG_FILE='master3-binlog.000001', MASTER_LOG_POS = 0;START SLAVE;"
mysql -h127.0.0.1 -P 3301 -e "RESET SLAVE; RESET MASTER; SHOW MASTER STATUS\G;"
mysql -h127.0.0.1 -P 3301 -e "CHANGE MASTER TO MASTER_HOST='127.0.0.1', MASTER_PORT=3304, MASTER_USER='root', MASTER_PASSWORD='', MASTER_LOG_FILE='master4-binlog.000001', MASTER_LOG_POS = 0;START SLAVE;"
```




MultiSource?



М.ВидеоЭльдорадо



MultiSource

master

slave

slave

slave

master2

master3

master1

master4

slave

CHANGE MASTER TO master_log_file='mysql-bin.0001', master_log_pos=0 **FOR CHANNEL “master1”;**
SHOW SLAVE STATUS **FOR CHANNEL “master2”;**

М.ВидеоЭльдорадо



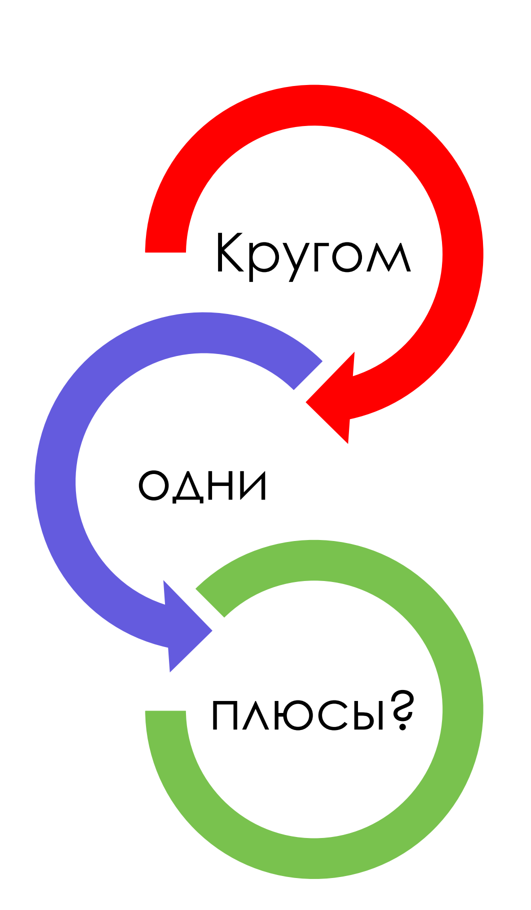
Новая схема

slave-db1-1
slave-db2-1
slave-db3-1
slave-db4-1
master01
master03
master04
master02
slave-db1-2
slave-db2-2
slave-db3-2
slave-db4-2
slave-db{2,3,4}-100
repl-to-dev

М.ВидеоЭльдорадо



План миграции



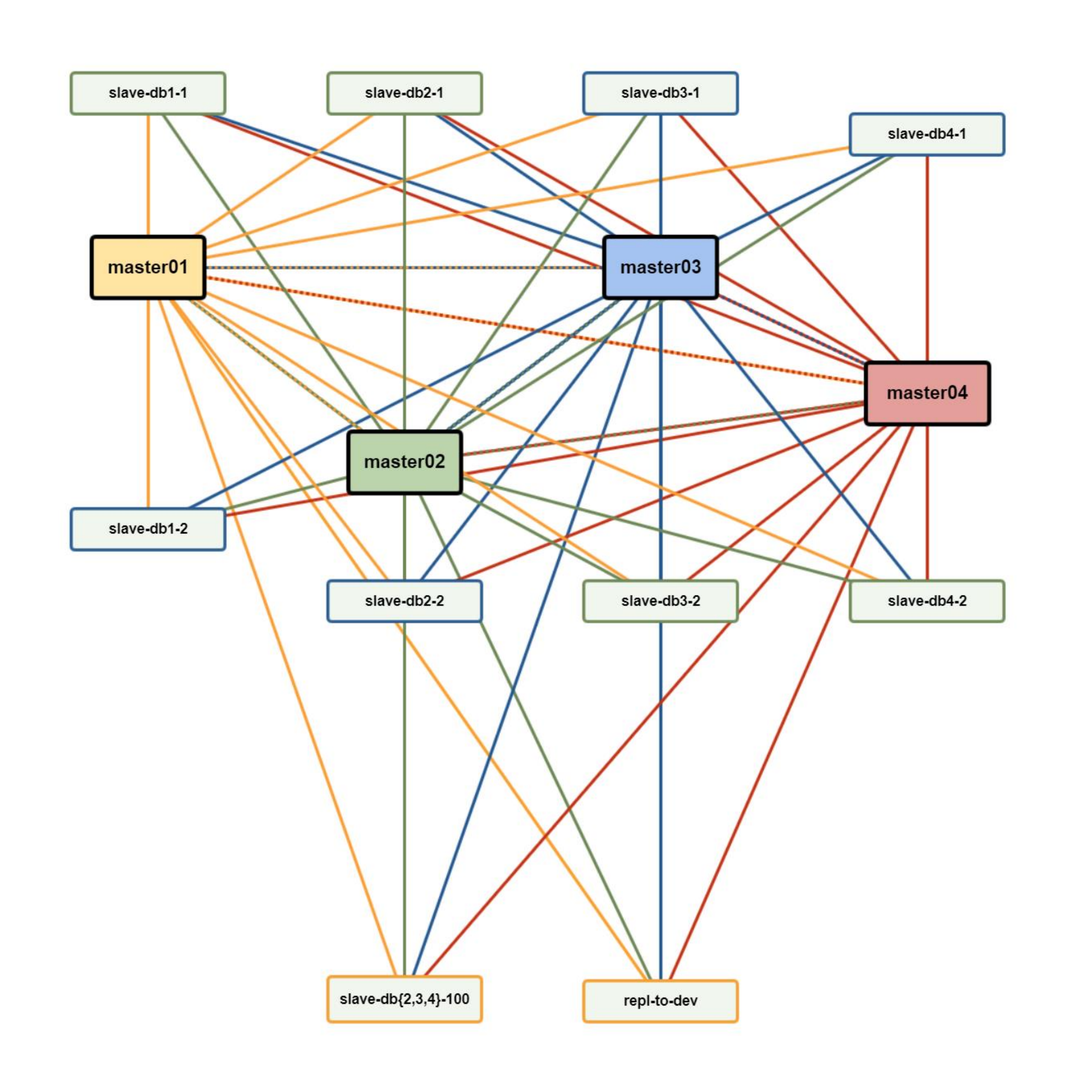
slave-db1-1
slave-db4-1
slave-db1-2
master01
master04
slave-db4-2
OLD
slave-db2-1
master02
master03
slave-db3-1
slave-db2-2
slave-db3-2

Slave-db2-100

Multimaster01 +
log_slave_updates=1
multimaster02
multimaster03
multimaster04



Нюансы

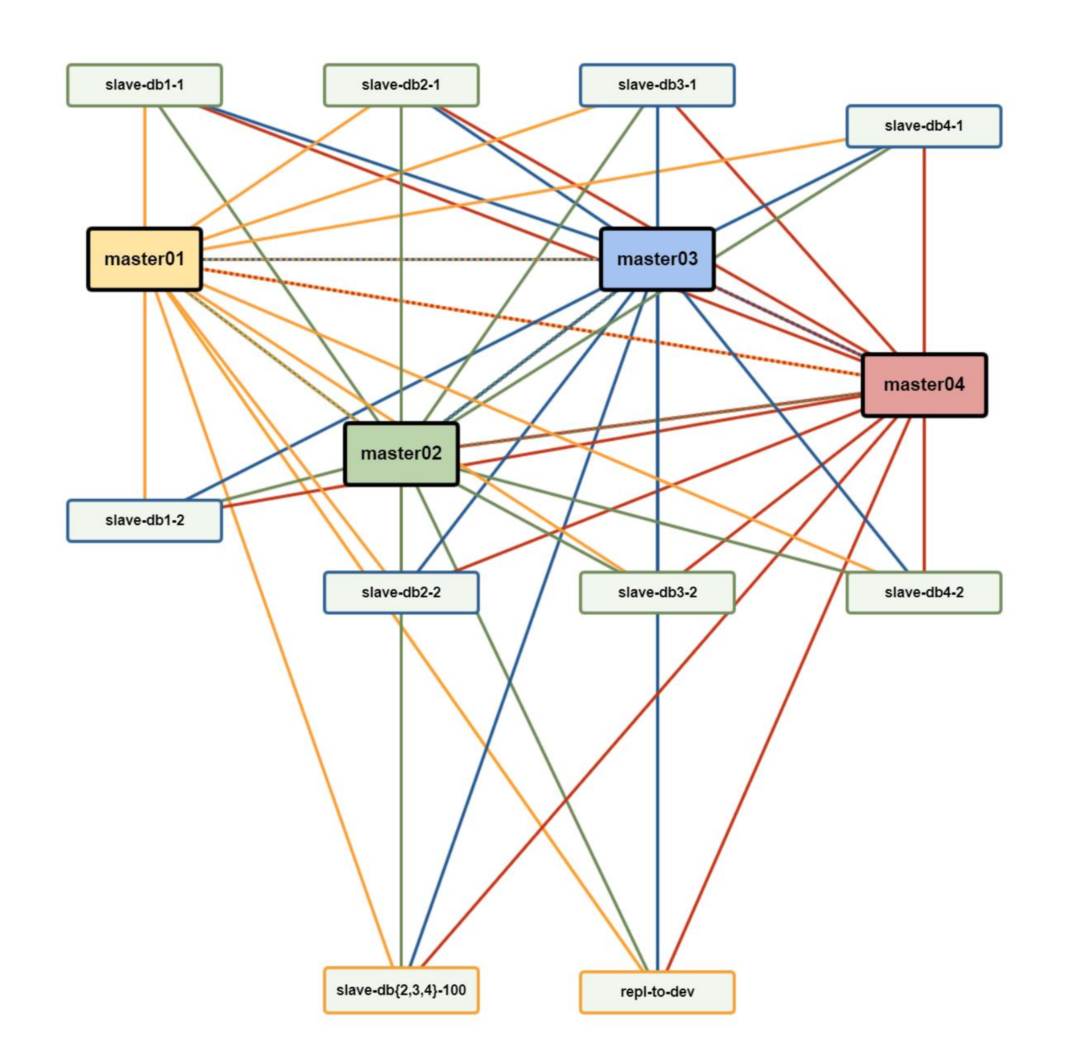


- Балансировка нагрузки
- Учесть возможный лаг репликации в приложении
- Инкрементные поля закончатся быстрее



Нюансы

- Не забыть про оффсеты
- Нагрузка на сеть на мастерах
- Онлайн миграции



binlog_row_image=minimal

М.ВидеоЭльдорадо



Online миграции

https://github.com/github/gh-ost



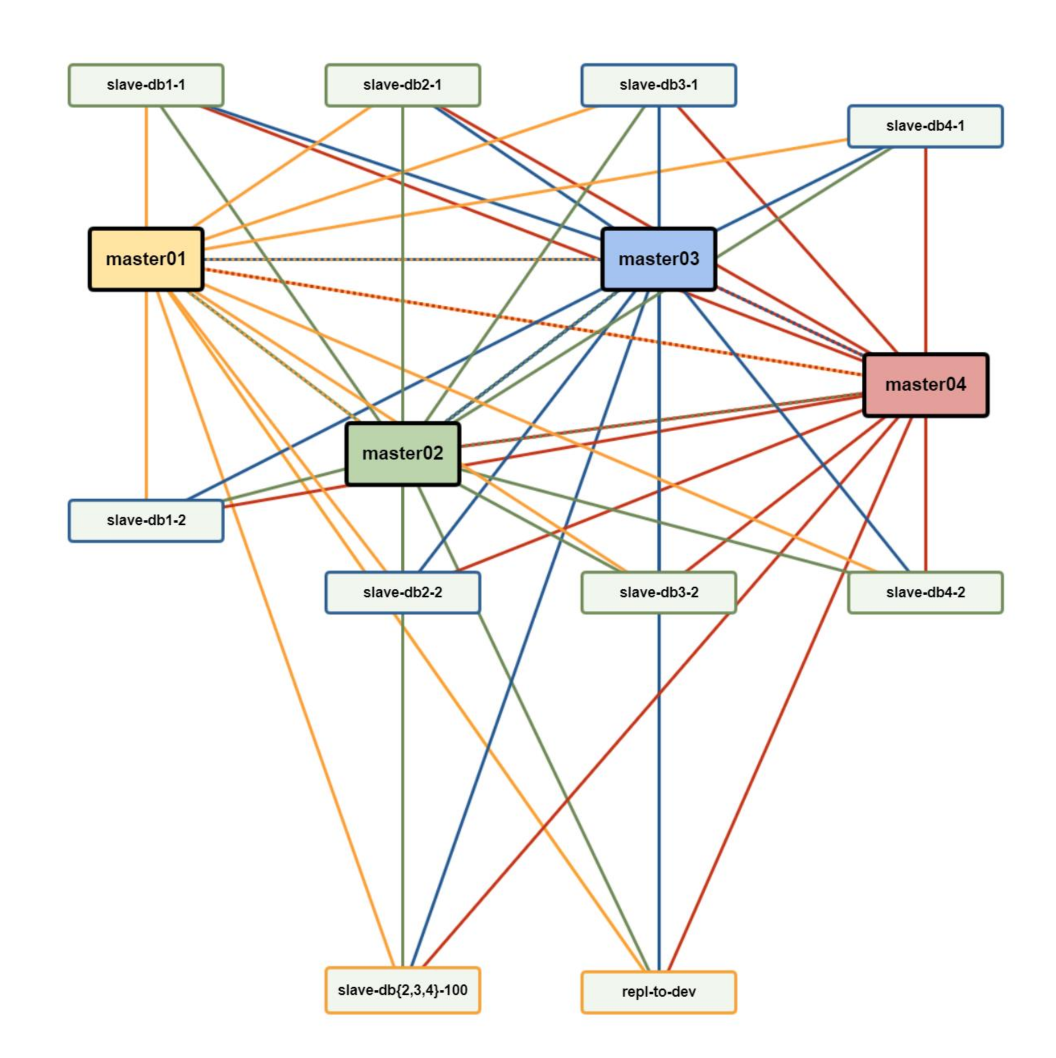
М.ВидеоЭльдорадо



Минусы

- Сложнее настраивать, поддерживать, мониторить
- Разная скорость проигрывания бинлогов
- Неконсистентность

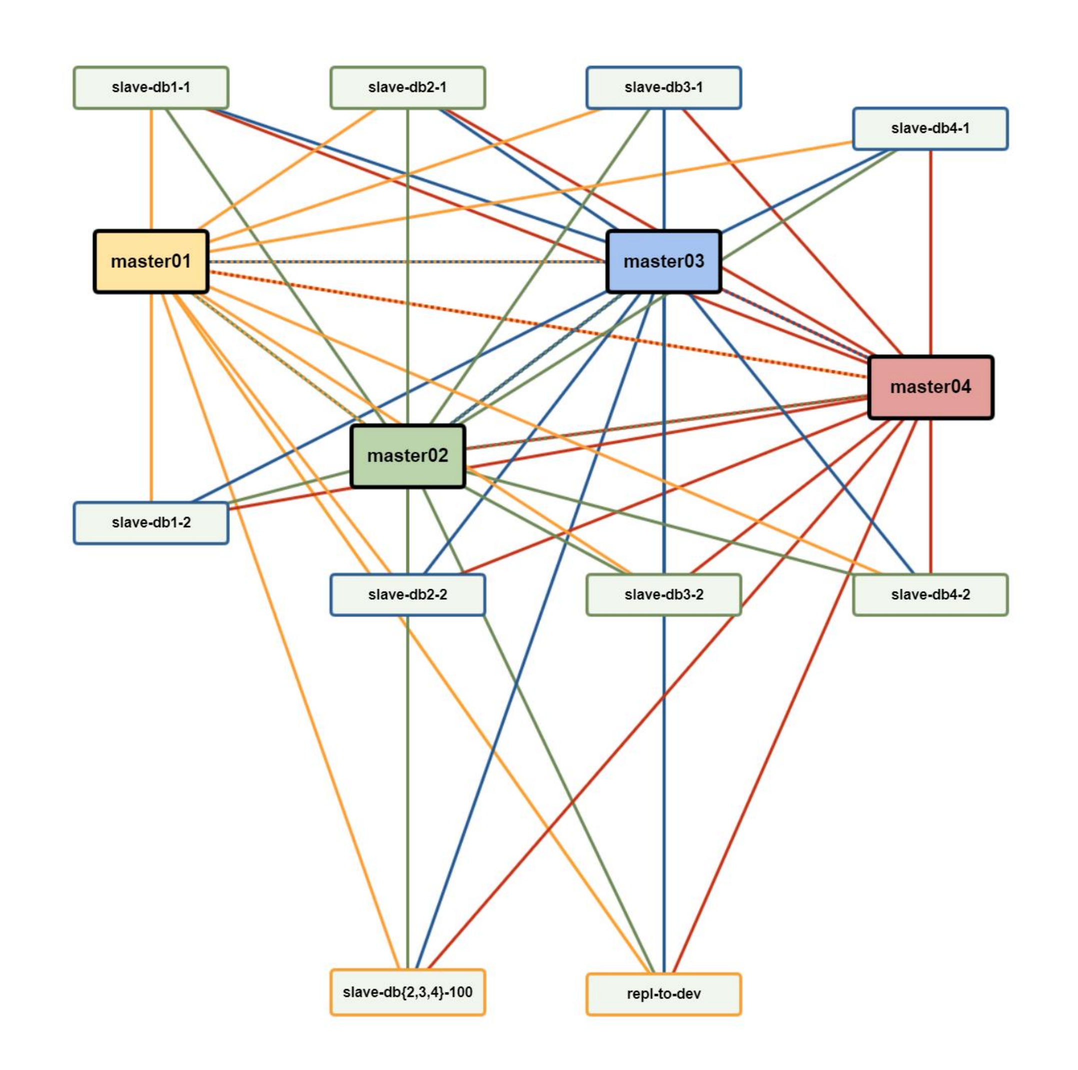
slave_parallel_workers=5
slave_parallel_type=LOGICAL_CLOCK



М.ВидеоЭльдорадо



Плюсы



- Легко переживает отказ любого мастера
- Меньше размер бинлогов
- Всегда только 1 плечо до любого слейва
- Лаги любого мастера влияют только на 25% потока записи

М.ВидеоЭльдорадо



Факап года

ради собственной
безопасности и
безопасности окружающих
не пытайтесь повторить

М.ВидеоЭльдорадо

Вопросы?

